J. Inv. Ill-Posed Problems **18** (2010), 701 – 726
DOI 10.1515/JIIP.2010.033

Conjugate gradient regularization under general smoothness and noise assumptions

Gilles Blanchard and Peter Mathé

Abstract. We study noisy linear operator equations in Hilbert space under a self-adjoint operator. Approximate solutions are sought by conjugate gradient type iteration, given as Krylov-subspace minimizers under a general weight function. Solution smoothness is given in terms of general source conditions. The noise may be controlled in stronger norm. We establish conditions under which stopping according to a modified discrepancy principle yields optimal regularization of the iteration. The present analysis extends much of the known theory and reveals some intrinsic features which are hidden when studying standard conjugate gradient type regularization under standard smoothness assumptions. In particular, under a non-self adjoint operator, regularization of the associated normal equation is a direct consequence from the main result and does not require a separate treatment.

Keywords. Conjugate gradient type regularization, general smoothness.

2010 Mathematics Subject Classification. 65J20; 65J22, 65F10.

1 Introduction, setup and main result

We shall analyze conjugate gradient (**cg**) type methods for solving (ill-posed) equations

$$y^{\delta} = Tx + \delta\xi, \tag{1.1}$$

for some bounded operator $T\colon X \to Y$ between Hilbert spaces, and noisy data y^{δ} with noise level $\delta > 0$. The assumptions on the noise ξ will be rather general, and we postpone discussion on this.

cg is originally designed for symmetric problems, i.e., when the operator governing the equation (1.1) is symmetric, self-adjoint and non-negative. To clearly distinguish this from the general case, we shall denote the corresponding operator by $A\colon X \to X$, and we thus consider the equation (in X) given as

$$z^{\delta} = Ax + \delta\eta. \tag{1.2}$$

The first author was holding a position at the Weierstrass Institut for Applied Analysis and Stochastics (Berlin) while carrying out this research.

The operator A need not be injective, and the solution to (1.2) can only be reconstructed for $x \in \ker^{\perp}(A)$, the orthogonal complement of the kernel of A. Since A is self-adjoint we have for the closure of the range that $\overline{\mathcal{R}(A)} = \ker^{\perp}(A)$, and we denote by Q the orthogonal projection onto $\overline{\mathcal{R}(A)}$.

The regularizing properties of the conjugate gradient iteration (**cg**) were first studied in the ground-breaking work [11]. Here we base our outline on [1, Chapter 7], and [2].

The classical conjugate gradient iteration (**cg**) minimizes the norm of the residual $z^{\delta} - Ax$ within corresponding *Krylov-subspaces*, i.e., the k-th iterate is given as

$$\|z^{\delta} - Ax_k^{\delta}\| = \min\{\|z^{\delta} - Ax\|,\ x \in \mathcal{K}_k(z^{\delta}, A)\},$$

where the Krylov subspace $K_k(z^{\delta}, A)$ consists of all elements $x \in X$ of the form $x = \sum_{j=0}^{k-1} c_j A^j z^{\delta}$. Thus the k-th **cg** iterate x_k^{δ} can be written as $x_k^{\delta} = g_k(A) z^{\delta}$, where g_k is a $(k-1)$-st order real polynomial.

1.1 Notation and convention for functional calculus

As a preliminary we recall very briefly the setting of functional calculus on operators and state the notational conventions which we follow in the rest of the paper. We let $\{E_{\lambda},\ 0 \le \lambda \le \|A\|\}$ be the spectral resolution of the self-adjoint semi-definite operator $A\colon X \to X$.

In the sequel we will at times consider piece-wise continuous functions f defined on $[0, \|A\|]$, taking real values in $(0, \|A\|]$ and whose value in 0 might be ∞. We will call such an f a *generalized in* 0 (in short g0) function, and denote f_+ its restriction on $(0, \|A\|]$. Given a g0 function f, assign the (possibly unbounded) operator $f(A)$ as

$$\langle f(A)x, z\rangle := \int_0^{\|A\|} f(\lambda)\, d\langle E_{\lambda}x, z\rangle \quad x, z \in X,$$

with domain

$$\mathcal{D}(f(A)) := \left\{x,\ \int_0^{\|A\|} f^2(\lambda) d\|E_{\lambda}x\|^2 < \infty\right\} \subset X.$$

If $f(0) = \infty$, then the latter integral is to be interpreted as being ∞ if x has a non-zero component on $\ker(A)$, and otherwise as the integral over $(0, \|A\|]$ of the same quantity; a similar convention applies for the first integral. The operator $f(A)$ is bounded if the function f is bounded, otherwise, the mapping $f(A)$ may give rise to an unbounded operator in X. By allowing the possibility of $f(0) = \infty$ we want to encompass different cases at once using the same notation, in particular whether or not a value of $f(A)$ is defined on $\ker(A)$ when f has a singularity in 0^+. In particular, a g0 function f will be said to be continuous in 0 if $f(0) = \lim_{t\to 0^+} f(t)$, encompassing both cases $f(0) < \infty$ and $f(0) = \infty$.

Observe that these refinements are only relevant for the case where A is not injective. The reader wanting to avoid these details can assume safely that A is injective ($\ker(A) = \{0\}$), and correspondingly skip discussions concerning the value of such functions at the origin.

To give a familiar example, assume $f(t) = t^{-1}$ for $t > 0$. If $f(0) = \infty$, then $f(A) = A^{-1}$, the usual inverse of A with domain $\mathcal{R}(A)$; if $f(0) = 0$, then $f(A) = A^+$, the Moore–Penrose inverse of A, which coincides with the usual inverse on its domain and is extended to take the value 0 on $\ker(A)$. In general, $f(0) < \infty \Leftrightarrow \ker^\perp(A) \subset \mathcal{D}(f(A))$.

1.2 General conjugate gradient type iteration

The classical **cg** approach extends to more general *conjugate gradient type iteration*, and we consider **cg**(w), for an arbitrary weight function $w\colon [0, \|A\|] \to \mathbb{R}^+$, such that w_+ is continuous and strictly positive. Given a weight w, for which $\|w(A)z^\delta\| < \infty$, conjugate gradient type iterations are derived from the following minimization in a Krylov subspace:

$$\|z^\delta - Ax_k^\delta\| = \min\{\|w(A)(z^\delta - Ax)\|, x \in \mathcal{K}_k(z^\delta, A)\}, \tag{1.3}$$

Let us denote, as above, by

$$x_k^\delta := g_k(A)z^\delta, \quad k = 1, 2 \ldots, \tag{1.4}$$

the corresponding minimizer in (1.3), where we recall that, by construction, $g_k(A)$ is a $(k-1)$-st order polynomial. We also denote $x_0^\delta = 0$ and $g_0 = 0$. Since $\|w(A)(z^\delta - Ax_k^\delta)\| < \infty$ then

$$w(A)(z^\delta - Ax_k^\delta) = w(A)r_k(A)z^\delta = r_k(A)\, w(A)z^\delta, \tag{1.5}$$

where we denote by $r_k(\lambda) := 1 - \lambda g_k(\lambda)$, the residual polynomial of degree k.

Remark 1.1. The classical **cg** corresponds to the weight $w_0(\lambda) = 1$. The value of the weight at zero, $w(0)$, does not effect the minimization procedure, since if $w(0) < \infty$ we have

$$\|w(A)(z^\delta - Ax)\|^2 = \|w(A)(Qz^\delta - Ax)\|^2 + w(0)^2\|(I - Q)z^\delta\|^2.$$

The last summand is constant, and hence the minimizer in (1.3) is the same as the one of $\|w(A)(Qz^\delta - Ax)\|$. To sum up, only w_+ matters for the definition of the **cg** iterations, and without loss of generality we can assume $w(0) = 0$.

Observe that the generated polynomials are different for different data z^δ, which makes **cg**(w) a non-linear iteration, and complicates the analysis.

Remark 1.2. If the data z^δ are degenerate in the sense that the measure $d\,\|E_\lambda z^\delta\|^2$ has only a finite number κ of non-zero points of increase, then for $k > \kappa$ the residuals $\|z^\delta - Ax_k^\delta\|$ remain constant, and the polynomials g_k are not uniquely defined from iteration κ on. We shall carry out the minimization along with some stopping criterion, a version of the discrepancy principle, and it will be clear from the discussions in Remarks 1.7 and 3.1 that the procedure will stop before iteration κ. If this degenerate situation does not present itself, we define $\kappa = \infty$.

1.3 General smoothness

We consider the following general notion of smoothness. We assume that there is an increasing continuous function $\psi\colon [0, \|A\|] \to \mathbb{R}^+,\ \psi(0) = 0$, for which

$$x \in H_\psi := \{\psi(A)v,\ \|v\| \le 1\} \subset \ker^\perp(A), \tag{1.6}$$

the image of the unit ball in X under the mapping $\psi(A)$; the set H_ψ is dense in $\ker^\perp(A)$.

Remark 1.3. A crucial observation, made recently in [5], asserts that each element $x \in \ker^\perp(A)$ has a certain amount of smoothness, meaning that there exists a function ψ satisfying the above conditions and such that $x \in H_\psi$. Low smoothness corresponds to concavity of the function ψ, and it was shown in [6] that such a choice is always possible. This fact will gain importance later.

For the later analysis we agree upon the following notion.

Definition 1.4. Let $f, g\colon (0, \|A\|] \to \mathbb{R}^+$ be two positive functions. The function f is said to be *majorized by* g if the function g/f is non-decreasing. We shall denote this by[1] $f \prec g$, or (with some abuse of notation) $f(\lambda) \prec g(\lambda)$. Specifically, in case that $g(\lambda) := \lambda^\mu$ we say that f is *majorized by the power* μ ($f(\lambda) \prec \lambda^\mu$). A function is said to be *majorized by a power* if it is majorized by the power μ for some $\mu > 0$.

Typically, if f and g are g0 functions, we will consider this type of relation for their restrictions f_+, g_+ on $(0, \|A\|]$.

[1] This defines a partial ordering. For monomials $f(\lambda) = \lambda^\mu,\ g(\lambda) := \lambda^\nu$ we have $f \prec g$ if and only if $\mu \le \nu$. Thus, the faster the function decays to zero (as $\lambda \to 0$) the more it is majorizing.

It is easy to see that if f is a concave function defined on $[0, \|A\|]$, with $f(0) \geq 0$, then f_+ is majorized by the power one ($f_+(\lambda) \prec \lambda^1$). Further relevant properties of such functions are postponed to Section 3. We stress however here the important property that if f, g are g0 functions, $f_+ \prec g_+$ implies $(\mathcal{D}(f(A)) \cap \ker^\perp(A)) \subset (\mathcal{D}(g(A)) \cap \ker^\perp(A))$, since for any $x \in \ker^\perp(A)$:

$$\|g(A)x\|^2 = \int_{0+}^{\|A\|} g^2(\lambda) d\,\|E_\lambda x\|^2 \leq \left(\frac{g(\|A\|)}{f(\|A\|)}\right)^2 \int_{0+}^{\|A\|} f^2(\lambda) d\,\|E_\lambda x\|^2$$
$$= \left(\frac{g(\|A\|)}{f(\|A\|)}\right)^2 \|f(A)x\|^2 .$$

Finally, throughout the study, we assign to any g0 function f the related real function

$$\Theta_f(\lambda) := \lambda f(\lambda), \quad 0 < \lambda \leq \|A\|. \tag{1.7}$$

Observe that the value of f in zero does not change the above definition, which effectively only depends on f_+.

1.4 General noise assumption

We impose the following general assumption concerning the noise under the model (1.2).

Assumption 1. There is a (bounded or unbounded) closed operator such that

$$\|L(z^\delta - Ax)\| \leq \delta. \tag{1.8}$$

For the subsequent analysis we need to relate the above operator L to the operator A governing the equation (1.2), using a function ϱ, and we introduce the following set of assumptions.

Assumption 2. The g0 function ϱ is such that its restriction function ϱ_+ on $(0, \|A\|] \to \mathbb{R}^+$ is non-negative, non-increasing, continuous, ϱ is continuous in zero, and the related function Θ_ϱ from (1.7) is non-decreasing.

Using the majorization from Definition 1.4 we can reformulate the above assumption as $\lambda^{-1} \prec \varrho_+(\lambda) \prec 1$ and ϱ_+ continuous. This is accompanied with the pair of complementary link conditions.

Assumption 3.

$$\|\varrho(A)z\| \leq \|Lz\|, \quad z \in \mathcal{D}(L). \tag{1.9}$$

In view of condition (1.8), and under Assumption 3 we have that

$$\|\varrho(A)(z^\delta - Ax)\| \leq \delta.$$

We mention the following instances. If $\varrho(0) = \infty$, then necessarily we have that $z^\delta \in \ker^\perp(A)$. If on the other hand $\varrho(0) = \varrho_0 < \infty$, then it is easy to see that $\varrho_0 > 0$ (by monotonicity, strict positivity and continuity in 0), and

$$\delta^2 \geq \|\varrho(A)(z^\delta - Ax)\|^2 = \varrho_0^2\|(I - Q)z^\delta\|^2 + \|\varrho(A)(Qz^\delta - Ax)\|^2, \quad (1.10)$$

and hence $\varrho_0\|(I - Q)z^\delta\| \leq \delta$. Since $\varrho_0 > 0$, this means that the size of the noise component in the kernel is limited.

Such functions ϱ from above are not unique. For instance, in the classical case when $L = I_X$ any non-increasing continuous and bounded (by one) ϱ will do. Thus, the 'maximal' functions will be of interest. These need not exist in general, but the next assumption will ensure that we are in this situation, and the convergence rates will depend on such ϱ.

Assumption 4. There is a constant $0 < m \leq 1$ for which

$$m\|Lz\| \leq \|\varrho(A)z\|, \quad z \in \mathcal{D}(\varrho(A)). \quad (1.11)$$

Remark 1.5. Assumption 1 implies in particular that $z^\delta - Ax \in \mathcal{D}(L)$. The assumption $\varrho(\lambda) \succ \lambda^{-1}$ (Assumption 2) implies $Ax \in \mathcal{D}(A^{-1}) \subset \mathcal{D}(\varrho(A))$. Assumption 3 implies $\mathcal{D}(L) \subset \mathcal{D}(\varrho(A))$. So, Assumptions 1–3 taken together imply $z^\delta \in \mathcal{D}(\varrho(A))$ and thus also $x_k^\delta \in \mathcal{D}(\varrho(A))$. Additionally, Assumption 4 implies the converse $\mathcal{D}(\varrho(A)) \subset \mathcal{D}(L)$.

1.5 The stopping criterion

Here we shall consider the following modification of the usual discrepancy principle to the present noise situation. Let L be given as in (1.8).

Definition 1.6. Given $\tau > 1$ we let the discrepancy parameter choice k_{DP} be the smallest $k \geq 0$ for which

$$\|L(z^\delta - Ax_k^\delta)\| \leq \tau\delta.$$

Remark 1.7. First, under Assumptions 1–4, and if $m\tau > 1$ then the number k_{DP} is well-defined: To this end let κ denote the index until which the minimization steps from (1.3) are well-defined, see the discussion in Remark 3.1. If $\kappa < \infty$ then we have that $Qz^\delta = Ax_\kappa^\delta$, and hence that

$$\|L(z^\delta - Ax_\kappa^\delta)\| = \|L(I - Q)z^\delta\| \leq \frac{1}{m}\|\varrho(A)(I - Q)z^\delta\|.$$

If $(I - Q)z^\delta = 0$ then $k_{\mathrm{DP}} \le \kappa$. In the remaining case, when $\varrho_0 < \infty$ and $(I - Q)z^\delta \neq 0$ then by (1.10) we have

$$\|\varrho(A)(I - Q)z^\delta\| = \frac{\varrho_0}{m}\|(I - Q)z^\delta\| \le \frac{\delta}{m} < \tau\delta.$$

Thus, if the number κ from Remark 3.1 is finite, then $k_{\mathrm{DP}} \le \kappa$. Otherwise, if $\kappa = \infty$, then this can be seen from Lemma 4.1, and the fact that $\alpha_k \searrow 0$ as $k \to \infty$.

Since we have to assume that $m\tau > 2$ below, we need to know a lower bound for the constant m in order to choose τ suitably. This poses no problem if L is an explicit function of A, as e.g., for the classical case $L = I$, since there $m = 1$.

1.6 Main results

From the discussion in §1.2–1.4 we have three functions which are important for the subsequent analysis, the weight w describing the conjugate gradient type iteration, ϱ controlling the noise, and ψ for the solution smoothness.

We state the following result on order optimal reconstruction, and we refer to [7]. To this end we consider the modulus of continuity of the Moore–Penrose inverse A^+, given as

$$\omega_\varrho(A^+, M, \delta) := \sup\{\|x\|,\ x \in M,\ \|\varrho(A)Ax\| \le \delta\}, \quad \delta > 0. \tag{1.12}$$

Furthermore, $\Theta_{\varrho\psi}$ is defined according to (1.7), using $f := \varrho\psi$. Since ϱ_+ and ψ are continuous, so is $\Theta_{\varrho\psi}$. We have $\Theta_{\varrho\psi} = \psi\Theta_\varrho$; since $t^{-1} \prec \varrho_+(t)$, Θ_ϱ is non-negative and non-decreasing and has therefore a non-negative limit in 0^+. Since ψ is continuous, increasing with $\psi(0) = 0$, we conclude that $\Theta_{\varrho\psi}$ is also continuous strictly increasing, with limit 0 in 0^+, and therefore has a well-defined inverse $\Theta_{\varrho\psi}^{-1}$ enjoying these same properties.

Fact 1. Suppose that the noise is controlled as in (1.8), and that Assumptions 1–4 hold. The best possible reconstruction accuracy valid uniformly over H_ψ equals $\omega_\varrho(A^+, H_\psi, \delta)$ and this obeys

$$\omega_\varrho(A^+, H_\psi, \delta) \asymp \psi(\Theta_{\varrho\psi}^{-1}(\delta)), \quad \delta \to 0,$$

provided that the function ψ is majorized by some power.

Thus, the question we tackle below is whether **cg**, or **cg**(w), stopped appropriately, can provide this optimal order of reconstruction.

Main Result. *Consider* x_k^δ *from* **cg**(w). *Suppose Assumptions* 1–4 *hold and assume* $\varrho_+ \prec w_+$. *Let* τ, $m\tau > 2$, *be a numerical constant and* k_{DP} *be the iteration picked by the discrepancy principle from Definition* 1.6.

If the solution $x \in H_\psi$ *for a function* ψ, *such that* $\psi_+(\lambda) \prec \lambda^\mu$, *then there is a constant* $C = C(\mu, m, \tau) < \infty$ *such that for* $x_{k_{\mathrm{DP}}}^\delta$ *we have that*

$$\|x - x_{k_{\mathrm{DP}}}^\delta\| \leq C(\mu, m, \tau)\psi(\Theta_{\varrho\psi}^{-1}(\delta)).$$

For the sake of coherence, let us check that all quantities in the above result are well-defined. Namely, under Assumptions 1–4, it holds that $z^\delta \in \mathcal{D}(\varrho(A))$ (see Remark 1.5), and thus also $z^\delta \in \mathcal{D}(w(A))$ under the assumption $\varrho_+ \prec w_+$. Hence both the algorithm and the stopping criterion are well-defined.

We mention a few consequences of the above general result.

The first consequence goes back to [12] within the traditional setup.

Corollary 1.8. *Consider* x_k^δ *from* **cg**(w). *Suppose Assumptions* 1–4 *hold and assume* $\varrho_+ \prec w_+$. *Let* τ, $m\tau > 2$, *be a numerical constant and* k_{DP} *be the iteration picked by the discrepancy principle from Definition* 1.6.

For each $x \in \ker^\perp(A)$ *we have that*

$$\|x - x_{k_{\mathrm{DP}}}^\delta\| \to 0, \quad \text{as } \delta \to 0.$$

Proof. It was mentioned in Remark 1.3 that for each element there is a concave function ψ, $\psi(0) = 0$, such that $x \in H_\psi$. Such function is majorized by the power 1 ($\psi_+(\lambda) \prec \lambda^1$), see the end of §1.3. Therefore, we can use the main result. Since $\psi(\Theta_{\varrho\psi}^{-1}(\delta)) \to 0$ as $\delta \to 0$, this implies the result. □

The second consequence concerns the optimality properties of **cg**(w) for non-symmetric equations (1.1), by using the normal equations

$$T^* y^\delta = T^* T x + \delta T^* \xi. \tag{1.13}$$

We shall use **cg**(w) based on $A := T^*T$, the data $z^\delta = T^* y^\delta$, and the classical discrepancy principle, i.e., the smallest k for which $\|y^\delta - T x_k^\delta\| \leq \tau\delta$.

Corollary 1.9. *Let* y^δ *be given from* (1.1) *and assume the noise model* $\|y^\delta - Tx\| \leq \delta$. *Consider* x_k^δ *from* **cg**(w), *based on the operator* $A := T^*T$. *Let* $\tau > 2$, *be a numerical constant and* k_{DP} *be the iteration picked by the classical discrepancy principle. If the solution* $x \in H_\psi$ *for a function* ψ, *such that* $\psi_+ \prec \lambda^\mu$, *then there is a constant* $C = C(\mu, \tau) < \infty$ *such that for* $x_{k_{\mathrm{DP}}}^\delta$ *we have that*

$$\|x - x_{k_{\mathrm{DP}}}^\delta\| \leq C(\mu, \tau)\psi(\tilde{\Theta}_\psi^{-1}(\delta)),$$

where $\tilde{\Theta}_\psi(\lambda) := \sqrt{\lambda}\psi(\lambda)$, $0 < \lambda \leq \|T^*T\|$.

Proof. For the classical non-symmetric case from (1.1), the operator

$$U := (T^*T)^{-1/2}T^*\colon Y \to X$$

can be seen to be a (bounded linear partial) isometry (elementary operator in the sense of [3]). Denoting Q_Y the orthogonal projector on $\ker^\perp(T^*)$, we have for any $y \in Y$ the identity $\|Uy\| = \|Q_Y y\|$.

Denote $\delta_0 = \|(I - Q_Y)y^\delta\|$. We have

$$\begin{aligned}\delta^2 = \delta_0^2 + \|Q_Y(y^\delta - Tx)\|^2 &= \delta_0^2 + \|U(y^\delta - Tx)\|^2\\ &= \delta_0^2 + \|(T^*T)^{-1/2}T^*(y^\delta - Tx)\|^2\\ &= \delta_0^2 + \|L(z^\delta - Ax)\|^2,\end{aligned}$$

where we have posed $A := T^*T$, $L := A^{-1/2}$, and $z^\delta := T^*y^\delta$.

Introduce additionally $\tilde{\delta}$, defined as $\tilde{\delta}^2 := \delta^2 - \tau^{-2}\delta_0^2$. Since $\tau > 1$, from the above display we deduce that $\|L(z^\delta - Ax)\| \le \tilde{\delta}$. We shall use the main result with A, L, z^δ as defined above, $m = 1$, and the function $\varrho(x) = x^{-1/2}$. The stopping criterion used in the main result is the smallest k such that

$$\|L(z^\delta - Ax_k^\delta)\|^2 = \|Q_Y(y^\delta - Tx_k^\delta)\|^2 \le \tau^2\tilde{\delta}^2 = \tau^2\delta^2 - \|(I - Q_Y)y^\delta\|^2,$$

and is therefore equivalent to the classical discrepancy principle. The function $\tilde{\Theta}_\psi$ is exactly Θ_{ϱ_ψ}, and the conclusion of the main result yields (since $\tilde{\delta} \le \delta$)

$$\|x - x_{k_{\mathrm{DP}}}^\delta\| \le C(\mu, m, \tau)\psi(\tilde{\Theta}_\psi^{-1}(\tilde{\delta})) \le C(\mu, m, \tau)\psi(\tilde{\Theta}_\psi^{-1}(\delta)). \qquad \square$$

We postpone discussion of the above results to Section 2. However, in the special case that w as well as ψ are power functions, then the discussion in Remark 4.5 reveals that the statement of Corollary 1.9 holds for the extended range $\tau > 1$, which is the natural assumption in that context.

The outline of the material is as follows. We start with a discussion of the assumptions and related issues of interest in Section 2. We then present auxiliary results, which are essentially known and standard in the analysis of conjugate gradient type iteration, in Section 3. We turn to proving the main result in Section 4, by starting from the fundamental error decomposition, Proposition 4.3. The main result is then proved in §4.2.

2 Discussion

Here we shall discuss several issues concerned with the setup and/or with applications of the results to specific situations. We first provide discussion on the setup presented in §1.2–1.5.

2.1 General conjugate gradient type iterations

The introduction of **cg**(w) with a general weight function appears to be new. As mentioned in Remark 1.1, the classical **cg** is obtained with $w_0(\lambda) = 1$. More generally, Hanke [2] considers weight functions $w_n(\lambda) := \lambda^{(n-1)/2}$, which leads to conjugate gradient type iteration with parameter n. For integers n this leads to implementable **cg**(n) procedures. Otherwise, for general weights w the procedure **cg**(w) is well-defined, at least until the number κ from Remark 3.1, but may hardly be implemented. However, as the analysis reveals, we do not need any specific properties of the weights to derive the results, and therefore we find it convenient to take this general point of view.

2.2 General smoothness

The analysis of **cg**(w) under general smoothness seems to be given here for the first time. Previous results for **cg**(n), as presented in [1, 2] are special instances of the ones given here. In particular, the present approach allows to analyze **cg** for severely ill-posed problems, i.e., when smoothness is logarithmic. As a specific flavor of this general smoothness we were able to deduce Corollary 1.8 without any additional efforts.

We emphasize one specific point. The concept of majorization, as given in Definition 1.4 proves important. Below, property (3.11) will play the role of a qualification of regularization; it can only be established for monomials, and the majorization will allow us to extend this to other smoothness, see Lemma 3.5, below. This is no restriction for low smoothness; however, this excludes the consideration of mildly ill-posed problems when smoothness is exponential, as e.g. $\psi(\lambda) = e^{-a/\lambda}, \lambda > 0$ for some $a > 0$. The treatment of such type of ill-posed problems is scarce.

2.3 General noise

The setting from §1.4 is abstract. It can be seen from Assumptions 1–4, that regularization was achieved only for noise assumptions stronger than $L = I_X$, i.e., in general the operator L needs to be unbounded, and this is reflected in Assumption 2 by requiring that $\varrho_{-1} := \lambda^{-1} \prec \varrho \prec 1 =: \varrho_0$, resulting in the extremal cases

$L = I_X$: This is the usual error criterion for self-adjoint ill-posed problems in Hilbert space under bounded deterministic noise.

$L := A^{-1}$: This is a rather strong assumption, as not arbitrary data $z^\delta \in X$ are possible. Instead, necessarily we have that $z^\delta \in \mathcal{R}(A)$, and this yields that the reconstruction problem behaves well-posed.

The noise model for ϱ_0 is the ordinary one for self-adjoint operator equations. The noise model ϱ_{-1} requires that the observed data are smooth, i.e., $z^\delta \in \mathcal{R}(A)$, and this means that $z^\delta \in \mathcal{D}(A^+)$, the domain of the Moore–Penrose inverse. A look at the modulus of continuity from (1.12) reveals that for this noise model we have $\omega_{\varrho_{-1}}(A^+, H_\psi, \delta) \le \delta$, which corresponds to well-posedness of the problem, and the reconstruction $A^+ z^\delta$ is (up to a constant) optimal.

We emphasize that there is no regularization under "large noise" within the model (1.2). This specific situation occurs only for the non-symmetrized equation and it is a result of symmetrization, see §2.5 for details.

The regularizing properties of **cg**(w) (under the discrepancy principle) can only be established for weights w which are properly related to the operator L. We shall therefore give a brief motivation for the Assumptions 3 and 4, here. We start from (1.5), which formally can be rewritten as

$$\|w(A) r_k(A) z^\delta\| = \|w(A) L^{-1}\, L r_k(A) z^\delta\|,$$

under the noise assumption. Thus, if $\|w(A)L^{-1} \colon X \to X\|$ is bounded then we have that

$$\|w(A) r_k(A) z^\delta\| \le \|w(A)L^{-1} \colon X \to X\| \cdot \|L(z^\delta - A x_k^\delta)\|.$$

It is known that the boundedness of the composition $w(A)L^{-1}$ is equivalent to requiring that there is some constant $C < \infty$ for which

$$\|w(A)x\| \le C\|Lx\|, \quad x \in \mathcal{D}(L). \tag{2.1}$$

Thus Assumptions 3 and 4, together with the subsequent requirement that $\varrho \prec w$ will ensure that (2.1) holds. The iterates x_k^δ are the same if we multiply the weight w by a constant, since the orthogonal polynomials do not change. Therefore, it is natural to assume the constant $C = 1$ in Assumption 3. The complementary Assumption 4 allows to identify 'maximal' noise models ϱ.

2.4 General discrepancy principle

We used a generalization of the discrepancy principle, which suits the present noise model. Of course, for the classical case $L = I_X$ this coincides with the ordinary discrepancy principle.

The restriction for τ in the main result to ensure that $m\tau > 2$ seems artificial. Indeed, as will be discussed in Remark 4.5, in many cases this can be relaxed to the standard requirement that $m\tau > 1$. We formulated the more restrictive assumption in order to avoid additional requirements on ϱ and w.

We mention that other parameter choice can also be considered, as e.g. the Lepskiĭ balancing principle. However, this is not convenient for iterative regularization, since this starts from the last iteration step.

The use of other stopping criteria for **cg** type iterations under general noise assumptions is an open issue (results in this direction for linear regularization will be published in the forthcoming study [8]).

2.5 The non-symmetric equation

We already indicated in Corollary 1.9 why the classical case of a non-symmetric equation (1.1) under the noise model $\|y^\delta - Tx\| \leq \delta$ is a special instance of the present setup. This extends as follows. For non-symmetric equations (1.1) the noise model $\varrho_0(\lambda) = 1$ leads to the weak noise model

$$\|T^*(y^\delta - Tx_k^\delta)\| = \|z^\delta - T^*Tx_k^\delta\| \leq \delta. \tag{2.2}$$

Such a noise model was, to the best of our knowledge, first considered in [10], and called 'large noise' there. Thus, the standard error control assumption for the symmetric equation (1.2) corresponds to the error control under the weak noise model for the non-symmetric equation (1.1). While the latter situation was not explicitly considered before in studies concerning **cg**, through this correspondence existing results obtained for the symmetric equation under the standard noise model (for example in [2]) therefore imply it as a corollary.

2.6 Treating weaker noise

It was mentioned in §2.3 that for the symmetric equation, noise which is weaker than $z^\delta \in X$ cannot be treated. In the non-symmetric case, the extension to treat weak noise appears as a result of the symmetrization (1.13). Analogously, we could also apply (powers of) A to equation (1.2), resulting for instance in

$$Az^\delta = A^2x + \delta\eta.$$

By doing so, we could treat weaker noise in the symmetric case, requiring that $Az^\delta \in X$, only, but basing conjugate gradient type iterations on the operator A^2. This is a common approach, and for linear regularization this is called *general linear regularization*, sometimes. We refer to [13], and the recent treatment for weak noise in [8].

3 Auxiliary results

In the analysis we shall need a series of auxiliary results, many of those have their origin in the original monographs [1, 2]. However, to make this study easier accessible, we capture some of the crucial ideas.

3.1 Functions which are majorized by some power

We first elaborate of some convenient consequences of functions majorized by some power according to Definition 1.4. The restriction to classes of functions which are bounded by some power has technical reasons. Restricting the analysis to such class provides us with some convenient tools.

If $q > 1$ is any real number and if $\psi(\lambda) \prec \lambda^{\mu}$, then

$$\psi(q\lambda) \le q^{\mu}\psi(\lambda) \le q^{\mu}\psi(q\lambda), \quad 0 < \lambda \le \|A\|/q.$$

Functions with such bound control are often said to *obey a Δ_2-condition*. The class of such functions is closed under point-wise multiplications, in particular, if $\psi(\lambda) \prec \lambda^{\mu}$ then $\Theta_{\psi}(\lambda) \prec \lambda^{\mu+1}$. This class is however not invariant with respect to taking inverse, as shows the example $\psi(\lambda) = \log^{-1} 1/\lambda,\ 0 < \lambda < 1$.

Finally, with $\lambda = \Theta_{\psi}(s)$, the identity

$$\frac{\lambda}{\psi(\Theta_{\psi}^{-1}(\lambda))} = \frac{\Theta_{\psi}(s)}{\psi(s)} = s = \Theta_{\psi}^{-1}(\lambda)$$

shows that the composition obeys $\psi \circ \Theta_{\psi}^{-1}(\lambda) \prec \lambda^{1}$.

3.2 Residual polynomials and general calculus

From now on we fix some weight function w, and hence consider iterates according to **cg**(w), as introduced in Section 1. The major properties for the analysis of the conjugate gradient iteration are sketched as follows.

If the measure

$$dm(\lambda) := \lambda d\|E_{\lambda}w(A)z^{\delta}\|^2 = \lambda w^2(\lambda)d\|E_{\lambda}z^{\delta}\|^2 \tag{3.1}$$

is finite, i.e., $\int_0^{\|A\|} \lambda w^2(\lambda)d\|E_{\lambda}z^{\delta}\|^2 = \|w(A)A^{1/2}z^{\delta}\|^2 < \infty$ then there is a sequence $r_1, r_2, \ldots$ of orthogonal polynomials with respect to the measure m, each of them obeys $r_k(0) = 1$, and r_k has degree equal to k. Such polynomials can be represented as $r_k(\lambda) = 1 - \lambda g_k(\lambda)$. These are the functions g_k used to assign x_k^{δ} as in (1.4).

Under the more restrictive assumption $\|w(A)z^\delta\| < \infty$ the orthogonal polynomials r_k enjoy the following optimality property: For each k, the orthogonal polynomial r_k, which realizes the minimum in (1.3) is minimal among all polynomials $\varphi \in \Pi_k$ of degree k with $\varphi(0) = 1$, i.e., we have that

$$\|r_k(A)w(A)z^\delta\| \le \|\varphi(A)w(A)z^\delta\|, \quad \varphi \in \Pi_k,\ \varphi(0) = 1.$$

Remark 3.1. The sequence $r_1, r_2, \dots, r_k$, of orthogonal polynomials is well-defined for k smaller than the number κ of points of increase of the measure m from (3.1), only. If $\kappa < \infty$ then we can use the discussion from [2, p. 11, after eq. (2.12)], applied to $n = 1$, and data $y := w(A)z^\delta$ to see that regardless of the weight w we have that $Qz^\delta = Ax_\kappa^\delta$.

The follow-up discussion, in particular in §3.3, is concerned with $k \le \kappa$, only.

We mention the following simple calculus which proves useful in the analysis. If $f \in L_2(\mathbb{R}, d\|E_\lambda w(A)z\|^2)$ and if $z \in X$ then for $\varepsilon > 0$ and non-decreasing function ξ we have that

$$\int_\varepsilon^{\|A\|} f^2(\lambda)d\|E_\lambda w(A)z\|^2 \le \frac{1}{\xi^2(\varepsilon)} \int_\varepsilon^{\|A\|} f^2(\lambda)\xi^2(\lambda)d\|E_\lambda w(A)z\|^2,$$

which rewrites as

$$\|(I - E_\varepsilon)f(A)w(A)z\| \le \frac{1}{\xi(\varepsilon)}\|(I - E_\varepsilon)f(A)\xi(A)w(A)z\|.$$

Below, for any positive g0 function f we denote

$$\delta_k^{[f]} := \|f(A)(z^\delta - Ax_k^\delta)\|, \quad k = 1, 2 \dots, \tag{3.2}$$

the k-th f-discrepancy of x_k^δ under **cg**(w).

3.3 Properties of orthogonal polynomials on the real line

We turn to important properties of real orthogonal polynomials on the real line. For each $1 \le k \le \kappa$ the polynomial r_k has zeroes

$$0 < \lambda_{1,k} < \lambda_{2,k} < \dots < \lambda_{k,k} \le \|A\|,$$

hence r_k and its derivative satisfy

$$r_k(\lambda) = \prod_{j=1}^k \left(1 - \frac{\lambda}{\lambda_{j,k}}\right) \quad \text{and} \quad |r_k'(0)| = \sum_{j=1}^k \frac{1}{\lambda_{j,k}}, \quad k = 1, 2, \dots, \tag{3.3}$$

which implies $|r_k'(0)|^{-1} \leq \lambda_{1,k}$, and also $|r_k'(0)|^{-1} \leq \|A\|/k,\ k = 1, 2 \dots$. We mention that the zeroes of subsequent orthogonal polynomials are interlacing, see [1, Appendix A.2], and hence $|r_k'(0)| < |r_{k+1}'(0)|,\ k = 1, 2, \dots$. We will also use the convention $\lambda_{1,0} = \infty,\ r_0 \equiv 1$.

Moreover, on the interval $[0, \lambda_{1,k}]$ the polynomial r_k obeys that $0 \leq r_k(\lambda) \leq 1$, and it is decreasing and convex on this interval. The reciprocals of the quantities $|r_k'(0)|$ will be used throughout, and we agree to abbreviate

$$\alpha_k := |r_k'(0)|^{-1}, \quad k = 1, 2, \dots. \tag{3.4}$$

Along with the residual polynomials r_k we also consider the update polynomials (of degree $k-1$) for $1 \leq k \leq \kappa$:

$$u_k(\lambda) := \frac{r_{k-1}(\lambda) - r_k(\lambda)}{\lambda}, \quad \lambda \in \mathbb{R},$$

and the differences

$$\pi_k := r_{k-1}'(0) - r_k'(0) \geq 0. \tag{3.5}$$

Since the zeroes of r_k and r_{k-1} are interlacing it follows that the zeroes $\bar{\lambda}_{j,k-1}$, $j = 1, \dots, k-1$, of u_k interlace with the ones of r_{k-1} in the form

$$0 < \lambda_{1,k-1} < \bar{\lambda}_{1,k-1} < \cdots < \lambda_{k-1,k-1} < \bar{\lambda}_{k-1,k-1}.$$

Together with (3.5) this shows that

$$0 \leq u_k(\lambda) \leq \pi_k, \quad 0 < \lambda \leq \lambda_{1,k-1}. \tag{3.6}$$

The following identity is taken from [1, eq. (7.16)] (generalized, and in slightly modified notation), where $\delta_k^{[w]}$ is as in (3.2): For each weight w we have that

$$\|u_k(A)A^{1/2}w(A)z^\delta\|^2 = \pi_k[(\delta_{k-1}^{[w]})^2 - (\delta_k^{[w]})^2], \tag{3.7}$$

which, in turn, yields that

$$\|u_k(A)A^{1/2}w(A)z^\delta\| \leq \sqrt{\pi_k}\delta_{k-1}^{[w]}. \tag{3.8}$$

This allows to prove

Lemma 3.2. *For each $k \geq 1$ and $0 < \varepsilon \leq \lambda_{1,k-1}$ we have that*

$$\pi_k \delta_{k-1}^{[w]} \leq \pi_k \|E_\varepsilon w(A) z^\delta\| + \frac{1}{\sqrt{\varepsilon\pi_k}}\pi_k\delta_{k-1}^{[w]}. \tag{3.9}$$

Proof. Since $u_k(0)/\pi_k = 1$, and since the polynomials u_k have degree $k-1$, the optimality properties (1.3) of the polynomials r_k yield that

$$\begin{aligned}
\pi_k \delta_{k-1}^{[w]} &= \pi_k \|r_{k-1}(A) w(A) z^\delta\| \leq \|u_k(A) w(A) z^\delta\| \\
&\leq \|E_\varepsilon u_k(A) w(A) z^\delta\| + \|(I - E_\varepsilon) u_k(A) w(A) z^\delta\| \\
&\leq \pi_k \|E_\varepsilon w(A) z^\delta\| + \frac{1}{\sqrt{\varepsilon}} \|u_k(A) A^{1/2} w(A) z^\delta\| \\
&\leq \pi_k \|E_\varepsilon w(A) z^\delta\| + \frac{1}{\sqrt{\varepsilon \pi_k}} \pi_k \delta_{k-1}^{[w]},
\end{aligned}$$

where we used (3.6) and (3.8); this completes the proof of the lemma. □

3.4 Bounding the noise propagation

From the properties of orthogonal polynomials as outlined in §3.3, and with the above notation, we can derive the following technical result, which allows to control the noise propagation. We refer to [7] for general linear regularization.

Lemma 3.3. *Suppose that ϱ is a* g0 *function satisfying Assumption* 2*, i.e., ϱ_+ is a non-increasing continuous function on* $(0, \|A\|]$ *such that Θ_ϱ is non-decreasing, and ϱ is continuous in* 0 *in the generalized sense. Then*

$$\sup_{0 \leq \lambda \leq \lambda_{1,k}} |g_k(\lambda)| \frac{1}{\varrho(\lambda)} \leq \frac{1}{\Theta_\varrho(\alpha_k)}, \quad k = 1, 2, \ldots.$$

Proof. It follows from the convexity of the residual polynomials r_k on the interval $[0, \lambda_{1,k}]$ that

$$g_k(\lambda) = \frac{1 - r_k(\lambda)}{\lambda} \leq |r_k'(0)| = \frac{1}{\alpha_k}.$$

We distinguish between cases. If $0 < \lambda \leq \alpha_k$ then we use that $1/\varrho$ is non-decreasing, and we have that

$$|g_k(\lambda)| \frac{1}{\varrho(\lambda)} \leq \frac{1}{\alpha_k} \frac{1}{\varrho(\alpha_k)} = \frac{1}{\Theta_\varrho(\alpha_k)}.$$

If $\lambda = 0$, we take the limit in 0^+ of the above expression, using the continuity of ϱ in 0 (if $\varrho(0) = \infty$, this reasoning is still valid as $g_k(0) \neq 0$). Otherwise, if $\alpha_k \leq \lambda \leq \lambda_{1,k}$ then we use that $g_k(\lambda)\lambda = 1 - r_k(\lambda) \in [0, 1]$, and we bound

$$|g_k(\lambda)| \frac{1}{\varrho(\lambda)} = |g_k(\lambda)| \lambda \frac{1}{\lambda \varrho(\lambda)} \leq \frac{1}{\Theta_\varrho(\alpha_k)}.$$

This completes the proof. □

3.5 How cg(w) takes smoothness into account

Based on the above orthogonality relation the following functions prove important, namely

$$\varphi_k(\lambda) := r_k(\lambda)\left(\frac{\lambda_{1,k}}{\lambda_{1,k}-\lambda}\right)^{1/2}, \quad 0 < \lambda \le \lambda_{1,k},\ k \ge 1.$$

These resemble properties of qualification of linear regularization schemes, as will be evident from (3.11) and Lemma 3.5. The following estimate for $\delta_k^{[\varrho]}$ from (3.2), is fundamental, and can be seen as an extension of [2], eq. (3.8).

Lemma 3.4. *Let* x_k^δ *be obtained from* **cg**(w)*, with* $1 \le k \le \kappa$*. Let* ϱ *be a* g0 *function. If* $\varrho_+ \prec w_+$ *then*

$$\delta_k^{[\varrho]} = \|\varrho(A)(z^\delta - Ax_k^\delta)\| \le \|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)z^\delta\|. \tag{3.10}$$

Proof. We first decompose

$$\begin{aligned}\|\varrho(A)(z^\delta - Ax_k^\delta)\|^2 &= \|\varrho(A)r_k(A)z^\delta\|^2\\ &= \|Q\varrho(A)r_k(A)z^\delta\|^2 + \|(I-Q)\varrho(A)r_k(A)z^\delta\|^2\\ &= \varrho(0)^2\|Qz^\delta\|^2 + \|(I-Q)\varrho(A)r_k(A)z^\delta\|^2,\end{aligned}$$

where, in the case of $\varrho(0) = \infty$, the first term is to be interpreted as ∞ if $Qz^\delta \ne 0$ and 0 otherwise. Similarly,

$$\|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)z^\delta\|^2 = \varrho(0)^2\|Qz^\delta\|^2 + \|(I-Q)E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)z^\delta\|^2.$$

To establish the result, we are therefore reduced to compare the respective second terms of the above decompositions. Since $1 \le k \le \kappa$, the polynomial $\frac{r_k(\lambda)}{\lambda_{1,k}-\lambda}$ is of degree $k-1$ and is orthogonal to r_k with respect to the measure m defined in (3.1), implying

$$\begin{aligned}&\int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\frac{\lambda}{\lambda_{1,k}-\lambda}\, d\|E_\lambda w(A)z^\delta\|^2\\ &\quad= \int_{\lambda_{1,k}}^{\infty} r_k^2(\lambda)\frac{\lambda}{\lambda-\lambda_{1,k}}\, d\|E_\lambda w(A)z^\delta\|^2 \ge \int_{\lambda_{1,k}}^{\infty} r_k^2(\lambda)\, d\|E_\lambda w(A)z^\delta\|^2.\end{aligned}$$

This leads to the following bound:

$$\begin{aligned}\int_{\lambda_{1,k}}^{\infty} r_k^2(\lambda)\varrho^2(\lambda)d\|E_\lambda z^\delta\|^2 &\le \frac{\varrho^2(\lambda_{1,k})}{w^2(\lambda_{1,k})}\int_{\lambda_{1,k}}^{\infty} r_k^2(\lambda)d\|E_\lambda w(A)z^\delta\|^2\\ &\le \frac{\varrho^2(\lambda_{1,k})}{w^2(\lambda_{1,k})}\int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\frac{\lambda}{\lambda_{1,k}-\lambda}d\|E_\lambda w(A)z^\delta\|^2.\end{aligned}$$

We conclude that

$$
\begin{aligned}
&\|(I - Q)\varrho(A)r_k(A)z^\delta\|^2 \\
&= \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\varrho^2(\lambda)\, d\|E_\lambda z^\delta\|^2 + \int_{\lambda_{1,k}}^{\infty} r_k^2(\lambda)\varrho^2(\lambda)\, d\|E_\lambda z^\delta\|^2 \\
&\le \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\varrho^2(\lambda) d\|E_\lambda z^\delta\|^2 + \frac{\varrho^2(\lambda_{1,k})}{w^2(\lambda_{1,k})}\int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\frac{\lambda}{\lambda_{1,k}-\lambda} d\|E_\lambda w(A)z^\delta\|^2 \\
&\le \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\varrho^2(\lambda) d\|E_\lambda z^\delta\|^2 + \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\frac{\varrho^2(\lambda)}{w^2(\lambda)}\frac{\lambda}{\lambda_{1,k}-\lambda} d\|E_\lambda w(A)z^\delta\|^2 \\
&= \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\, d\|E_\lambda \varrho(A)z^\delta\|^2 + \int_{0+}^{\lambda_{1,k}} r_k^2(\lambda)\frac{\lambda}{\lambda_{1,k}-\lambda}\, d\|E_\lambda \varrho(A)z^\delta\|^2 \\
&= \|(I - Q)E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)z^\delta\|^2,
\end{aligned}
$$

where the latter equality uses the construction of the function φ_k. □

We return to properties of the functions φ_k. These function obey that $0 \le \varphi_k(\lambda) \le 1$, and, see [1, (7.8)], we have

$$\lambda^\nu \varphi_k^2(\lambda) \le \nu^\nu \alpha_k^\nu, \quad 0 < \lambda \le \lambda_{1,k}, \quad \nu > 0, \tag{3.11}$$

with α_k as in (3.4). Again, this has the following extension.

Lemma 3.5. *Let the positive non-decreasing function ξ be majorized by the power $\nu \ge 0$, i.e. $\xi(\lambda) \prec \lambda^\nu$. Then*

$$\varphi_k^2(\lambda)\xi(\lambda) \le \nu^\nu \xi(\alpha_k), \quad 0 < \lambda \le \lambda_{1,k}, \tag{3.12}$$

and, for any $v \in X$ with $\|v\| \le 1$ we have that

$$\|E_{\lambda_{1,k}}\varphi_k(A)A\xi(A)v\| \le (2\nu + 2)^{\nu+1}\Theta_\xi(\alpha_k).$$

Proof. For the first assertion we use the monotonicity of ξ to see that $\varphi_k^2(\lambda)\xi(\lambda) \le \xi(\alpha_k)$ for $\lambda \le \alpha_k$. Otherwise, if $\lambda > \alpha_k$ then we use (3.11) to bound

$$\varphi_k^2(\lambda)\xi(\lambda) = \varphi_k^2(\lambda)\lambda^\nu \frac{\xi(\lambda)}{\lambda^\nu} \le \nu^\nu {\alpha_k}^\nu \frac{\xi(\alpha_k)}{{\alpha_k}^\nu} = \nu^\nu \xi(\alpha_k),$$

which proves (3.12) in either case.

For the second estimate we rewrite

$$\|E_{\lambda_{1,k}}\varphi_k(A)A\xi(A)v\| = \|E_{\lambda_{1,k}}\varphi_k(A)\Theta_\xi(A)v\|.$$

Since ξ is majorized by the power ν, the function Θ_ξ is majorized by the power $\nu + 1$, and we apply (3.12) with the function $\xi' := \Theta_\xi^2 \prec \lambda^{2\nu+2}$. □

4 Error decomposition and proof of the main results

The steps to derive order optimal error bounds for **cg**(w) follow the development in [1, Chapter 7]. Thus, we shall find a suitable bound for the k-th ϱ-discrepancy from (3.2), which allows to provide an error decomposition, resembling the typical decomposition into regularization error (bias) and noise propagation (variance).

4.1 Error decomposition

We start with bounding the discrepancy for $x \in H_\psi$. Here, the function $\Theta_{\varrho\psi}$, see (1.7) with $f := \varrho\psi$, proves important.

Lemma 4.1. *Let x_k^δ be obtained from* **cg**(w). *Suppose Assumptions* 1–3 *hold. If $x \in H_\psi$, for a function ψ which is majorized by the power μ ($\psi(\lambda) \prec \lambda^\mu$) then we have*

$$\delta_k^{[\varrho]} = \|\varrho(A)(z^\delta - Ax_k^\delta)\| \le \delta + (2\mu + 2)^{\mu+1}\Theta_{\varrho\psi}(\alpha_k), \quad k = 1, 2, \dots.$$

Proof. We use Lemma 3.4; recall (see Remark 1.5) that by assumption on ϱ, $Ax \in \mathcal{D}(\varrho(A))$, allowing us to write

$$\begin{aligned} \|\varrho(A)(z^\delta - Ax_k^\delta)\| &\le \|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)(z^\delta - Ax)\| + \|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)Ax\| \\ &\le \delta + \|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)Ax\|, \end{aligned} \tag{4.1}$$

using that $\varphi_k(\lambda) \in [0, 1]$ for $\lambda \in [0, \lambda_{1,k}]$. Under $x \in H_\psi$ we have that $\varrho(A)Ax \in H_{\Theta_{\varrho\psi}}$. The function $\Theta_{\varrho\psi}$ is majorized by the power $\mu + 1$, and an application of Lemma 3.5 yields that

$$\|E_{\lambda_{1,k}}\varphi_k(A)\varrho(A)Ax\| \le (2\mu + 2)^{\mu+1}\Theta_{\varrho\psi}(\alpha_k).$$

The proof is complete. □

We continue and provide an intermediate error decomposition, using the k-th ϱ-discrepancy from (3.2).

Lemma 4.2. *Let $x_k^\delta = g_k(A)z^\delta$ be obtained from* **cg**(w), $1 \le k \le \kappa$, *and let $\lambda_{1,k}$ be the smallest zero of the corresponding polynomial r_k. Suppose Assumptions* 1–3 *hold.*

For $d_k := \max\{\delta, \delta_k^{[\varrho]}\}$, and for $0 < \varepsilon \le \lambda_{1,k}$ we have that

$$\|x - x_k^\delta\| \le \|E_\varepsilon x\| + \frac{2}{\Theta_\varrho(\varepsilon)}d_k + \frac{1}{\Theta_\varrho(\alpha_k)}\delta. \tag{4.2}$$

Proof. For any $\varepsilon > 0$ we can bound

$$\|x - x_k^\delta\| \leq \|E_\varepsilon(x - x_k^\delta)\| + \|(I - E_\varepsilon)(x - x_k^\delta)\|, \tag{4.3}$$

and we bound both summands, separately. The first summand from (4.3) is bounded as follows. Although the polynomial g_k is obtained at data z^δ we shall apply this formally to 'exact' data $z := Ax$, i.e., we let $\hat{x}_k := g_k(A; z^\delta)z$. With this notation we bound

$$\begin{aligned} \|E_\varepsilon(x - x_k^\delta)\| &\leq \|E_\varepsilon(x - \hat{x}_k)\| + \|E_\varepsilon(\hat{x}_k - x_k^\delta)\| \\ &= \|E_\varepsilon r_k(A; z^\delta)x\| + \|E_\varepsilon g_k(A; z^\delta)(z - z^\delta)\|. \end{aligned} \tag{4.4}$$

The first summand on the right above is bounded by $\|E_\varepsilon x\|$ for $0 < \varepsilon < \lambda_{1,k}$, and it remains to bound the last summand on the right. To this end we use Lemma 3.3 to conclude that

$$\begin{aligned} \|E_\varepsilon g_k(A; z^\delta)(z - z^\delta)\| &= \left(\int_0^\varepsilon \frac{g_k^2(\lambda; z^\delta)}{\varrho^2(\lambda)} \varrho^2(\lambda) d\|E_\lambda(z - z^\delta)\|^2 \right)^{\frac{1}{2}} \\ &\leq \frac{1}{\Theta_\varrho(\alpha_k)} \|\varrho(A)(z - z^\delta)\| \leq \frac{1}{\Theta_\varrho(\alpha_k)} \delta. \end{aligned}$$

The second summand on the right in (4.3) can be treated as

$$\begin{aligned} \|(I - E_\varepsilon)(x - x_k^\delta)\| &\leq \frac{1}{\Theta_\varrho(\varepsilon)} \|(I - E_\varepsilon)\varrho(A)(Ax - Ax_k^\delta)\| \\ &\leq \frac{2}{\Theta_\varrho(\varepsilon)} \max\{\delta, \delta_k^{[\varrho]}\}. \end{aligned}$$

Overall this results in the bound as stated. □

We turn to the major error decomposition, a consequence of Lemmas 4.1 and 4.2. Recall the sequence α_k from (3.4), and $\delta_k^{[\varrho]}$ from (3.10).

Proposition 4.3. *Consider x_k^δ from* **cg**(w)*, $1 \leq k \leq \kappa$, and suppose that Assumptions* 1–3 *holds true for a function ϱ, $\varrho_+ \prec w_+$. Let $d_k := \max\{\delta, \delta_k^{[\varrho]}\}$. If $x \in H_\psi$, for a function ψ, $\psi(\lambda) \prec \lambda^\mu$, then*

$$\|x - x_k^\delta\| \leq 2(1 + 2(2\mu + 2)^{\mu+1})\psi(\Theta_{\varrho\psi}^{-1}(d_k)) + \frac{3}{\Theta_\varrho(\alpha_k)}\delta.$$

Proof. First, for any positive increasing function ψ we have that $\psi(\lambda) \leq \psi(\varepsilon)$ whenever $0 < \lambda \leq \varepsilon$. Therefore (4.2) yields

$$\|x - x_k^\delta\| \leq \psi(\varepsilon) + \frac{2}{\Theta_\varrho(\varepsilon)} d_k + \frac{1}{\Theta_\varrho(\alpha_k)}\delta, \tag{4.5}$$

for $x \in H_\psi$ and $0 < \varepsilon < \lambda_{1,k}$. Let $\varepsilon_* := \Theta_{\varrho\psi}^{-1}(2d_k)$. To obtain an error bound we consider two cases.

If $\varepsilon_* < \alpha_k$ then we use (4.5) with $\varepsilon = \varepsilon_*$ to have

$$\|x - x_k^\delta\| \leq 2\psi(\Theta_{\varrho\psi}^{-1}(2d_k)) + \frac{1}{\Theta_\varrho(\alpha_k)}\delta.$$

Otherwise $\varepsilon_* \geq \alpha_k$, and we use (4.5) with $\varepsilon := \alpha_k$. Let us temporarily abbreviate $c(\mu) := (2\mu + 2)^{\mu+1}$. Lemma 4.1 allows to bound $\delta_k^{[\varrho]}$ and we have

$$\begin{aligned}
\|x - x_k^\delta\| &\leq \psi(\alpha_k) + \frac{2}{\Theta_\varrho(\alpha_k)} d_k + \frac{1}{\Theta_\varrho(\alpha_k)}\delta \\
&\leq \psi(\alpha_k) + \frac{2}{\Theta_\varrho(\alpha_k)}(\delta + c(\mu)\Theta_{\varrho\psi}(\alpha_k)) + \frac{1}{\Theta_\varrho(\alpha_k)}\delta \\
&\leq (1 + 2c(\mu))\psi(\alpha_k) + \frac{3}{\Theta_\varrho(\alpha_k)}\delta \\
&\leq (1 + 2c(\mu))\psi(\Theta_{\varrho\psi}^{-1}(2d_k)) + \frac{3}{\Theta_\varrho(\alpha_k)}\delta \\
&\leq 2(1 + 2c(\mu))\psi(\Theta_{\varrho\psi}^{-1}(d_k)) + \frac{3}{\Theta_\varrho(\alpha_k)}\delta,
\end{aligned}$$

since $\psi \circ \Theta_{\varrho\psi}^{-1}(\lambda) \prec \lambda^1$, see §3.1. The proof is complete. □

4.2 Proof of the main result

By using the weak discrepancy principle from Definition 1.6, and under Assumption 3 we get a suitable bound for $\delta_k^{[\varrho]}$ from above by its very definition, and this already provides us with the correct order of the first term in the error decomposition from Proposition 4.3, since then $d_{k_{\mathrm{DP}}} \leq \tau\delta$. To establish a bound for the second term, we need a lower bound for $\alpha_{k_{\mathrm{DP}}}$, and hence an upper bound for $|r'_{k_{\mathrm{DP}}}(0)|$. This will be accomplished by using Assumption 4 in addition to Assumption 3.

Lemma 4.4. *Let $x_k^\delta = g_k(A)z^\delta$ be obtained from* **cg**(w)*, and let k_{DP} be according to the weak discrepancy principle from Definition* 1.6 *with $\tau > 1$. Suppose Assumptions* 1–4 *hold. If the function ϱ is majorized by the weight w ($\varrho_+ \prec w_+$),*

$x \in H_\psi$ for a function ψ, $\psi(\lambda) \prec \lambda^\mu$, and if $m\tau > 2$ then there are two constants $1 \le C(\mu, m, \tau) < \infty$ and $0 < \zeta(\mu, m, \tau) \le 1$ such that

$$|r'_{k_{\mathrm{DP}}}(0)| \le C(\mu, m, \tau) \frac{1}{\Theta_{\varrho\psi}^{-1}(\zeta(\mu, m, \tau)\delta)}. \tag{4.6}$$

Proof. If $k_{\mathrm{DP}} = 0$, then $|r'_{k_{\mathrm{DP}}}(0)| = 0$ and there is nothing to prove. From now on assume $k_{\mathrm{DP}} \ge 1$. Using π_k from (3.5) we have that

$$|r'_{k_{\mathrm{DP}}}(0)| \le \pi_{k_{\mathrm{DP}}} + |r'_{k_{\mathrm{DP}}-1}(0)|.$$

By the definition of k_{DP} and by Lemma 4.1 we find that

$$\frac{m\tau - 1}{(2\mu + 2)^{\mu+1}} \delta \le \Theta_{\varrho\psi}(\alpha_{k_{\mathrm{DP}}-1}).$$

We temporarily abbreviate $\bar{c} := (m\tau - 1)/(2\mu + 2)^{\mu+1}$, and thus find that

$$|r'_{k_{\mathrm{DP}}-1}(0)| \le \frac{1}{\Theta_{\varrho\psi}^{-1}(\bar{c}\delta)}. \tag{4.7}$$

Let $\eta, \xi \le 1$ be positive factors which will be fixed later and can depend only on m, τ and μ, and such that $m\tau - 1 - \eta\bar{c} > 0$. The announced result will be established by taking $\xi = 1$, but we introduce slightly more generality here for the purpose of a discussion coming after the proof. Put $\varepsilon_* := \Theta_{\varrho\psi}^{-1}(\eta\bar{c}\delta)$. By virtue of (4.7) we have that $\varepsilon_* \le \lambda_{1,k_{\mathrm{DP}}-1}$. We want to establish that $\pi_{k_{\mathrm{DP}}}$ is of the order $1/\varepsilon_*$, and we shall use Lemma 3.2. Taking into account that ϱ is majorized by the weight w we have that

$$\begin{aligned} \|E_{\xi\varepsilon_*} w(A) z^\delta\| &= \left(\int_{0+}^{\xi\varepsilon_*} \frac{w^2(\lambda)}{\varrho^2(\lambda)} \varrho^2(\lambda) d\|E_\lambda z^\delta\|^2 \right)^{\frac{1}{2}} && (4.8) \\ &\le \frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)} (\|\varrho(A)(z^\delta - Ax)\| + \|E_{\xi\varepsilon_*}\varrho(A)Ax\|) \\ &\le \frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)} (\delta + \Theta_{\varrho\psi}(\xi\varepsilon_*)) \\ &\le \frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)} (\delta + \Theta_{\varrho\psi}(\varepsilon_*)) \\ &= \frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)} (1 + \eta\bar{c})\delta. && (4.9) \end{aligned}$$

For every $\varepsilon < \lambda_{1,k}$ and k we have that

$$\begin{aligned}\delta_k^{[\varrho]} = \|\varrho(A)r_k(A)z^\delta\| &\le \|E_\varepsilon \varrho(A)r_k(A)z^\delta\| + \|(I - E_\varepsilon)\varrho(A)r_k(A)z^\delta\| \\ &\le \|E_\varepsilon \varrho(A)z^\delta\| + \frac{\varrho(\varepsilon)}{w(\varepsilon)}\delta_k^{[w]} \\ &\le \delta + \|E_\varepsilon \varrho(A)Ax\| + \frac{\varrho(\varepsilon)}{w(\varepsilon)}\delta_k^{[w]} \\ &\le \delta + \Theta_{\varrho\psi}(\varepsilon) + \frac{\varrho(\varepsilon)}{w(\varepsilon)}\delta_k^{[w]}.\end{aligned}$$

Letting $\varepsilon = \varepsilon_*$ and $k := k_{\mathrm{DP}} - 1$, and by Remark 1.7 this yields

$$m\tau\delta \le \delta_{k_{\mathrm{DP}}-1}^{[\varrho]} \le (1 + \eta\bar{c})\delta + \frac{\varrho(\varepsilon_*)}{w(\varepsilon_*)}\delta_{k_{\mathrm{DP}}-1}^{[w]}, \tag{4.10}$$

and consequently, that

$$\delta \le \frac{1}{m\tau - \eta\bar{c} - 1}\frac{\varrho(\varepsilon_*)}{w(\varepsilon_*)}\delta_{k_{\mathrm{DP}}-1}^{[w]}.$$

Combining with (4.9) gives

$$\|E_{\xi\varepsilon_*} w(A)z^\delta\| \le \frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)}\frac{\varrho(\varepsilon_*)}{w(\varepsilon_*)}\frac{1 + \eta\bar{c}}{m\tau - \eta\bar{c} - 1}\delta_{k_{\mathrm{DP}}-1}^{[w]}. \tag{4.11}$$

In the particular case $\xi = 1$, we obtain

$$\|E_{\varepsilon_*} w(A)z^\delta\| \le \frac{1 + \eta\bar{c}}{m\tau - \eta\bar{c} - 1}\delta_{k_{\mathrm{DP}}-1}^{[w]}. \tag{4.12}$$

Inserting this into (3.9) of Lemma 3.2 we see that

$$1 \le \frac{1 + \eta\bar{c}}{m\tau - 1 - \eta\bar{c}} + \frac{1}{\sqrt{\varepsilon_* \pi_{k_{\mathrm{DP}}}}}.$$

Assuming $m\tau > 2$, we can choose η small enough (depending only on τ, m, μ) so that $\frac{1+\eta\bar{c}}{m\tau-\eta\bar{c}-1} < 1$.

This gives, for a suitable constant $\hat{c} > 0$ that

$$\pi_{k_{\mathrm{DP}}} \le \hat{c}\frac{1}{\varepsilon_*} = \frac{\hat{c}}{\Theta_{\varrho\psi}^{-1}(\eta\bar{c}\delta)},$$

from which the proof can easily be completed. □

Remark 4.5. The requirement $m\tau > 2$ can be relaxed to the natural $m\tau > 1$ in many cases. First, if $w = \varrho$, then $\delta_k^{[w]} = c\delta_k^{[\varrho]}$. In this case we use (4.9) and the left-hand side bound in (4.10), and we let $\xi = 1$, to deduce that $\|E_{\varepsilon_*} w(A) z^\delta\| \leq \frac{1+\eta\bar{c}}{m\tau}\delta^{[w]}_{k_{\mathrm{DP}}-1}$. We can choose η small enough so that the factor is < 1.

Secondly, if there is a constant $\gamma > 0$ such that $\lambda^\gamma \prec \frac{w}{\varrho}(\lambda)$, then we have

$$\frac{w(\xi\varepsilon_*)}{\varrho(\xi\varepsilon_*)}\frac{\varrho(\varepsilon_*)}{w(\varepsilon_*)} \leq \xi^\gamma.$$

We can first choose η such that $m\tau - 1 - \eta\bar{c} > 0$. Then we choose ξ small enough (depending only on τ, m, μ and γ) such that the factor in (4.11) is < 1. This covers the monomial cases as treated in [2]. However, it is not clear what happens in the "intermediate" case, i.e., when w strictly majorizes ϱ but $\frac{w}{\varrho}$ is not minorized by a power.

Before turning to the order optimality of **cg**(w) under the weak discrepancy principle from Definition 1.6 we shall initially discuss the possibility of having small data.

Lemma 4.6. *If $\|\varrho(A)z^\delta\| \leq C\delta$ then the zero solution yields an optimal (up to a factor $2(1 + C)$) solution for any $x \in H_\psi$ with data z^δ.*

Specifically, we have that $\|x\| \leq 2(1 + C)\psi(\Theta^{-1}_{\varrho\psi}(\delta)),\ \delta > 0$.

Proof. We bound

$$\begin{aligned}\|x\| &\leq \|E_\varepsilon x\| + \|(I - E_\varepsilon)x\| \\ &\leq \|E_\varepsilon x\| + \frac{1}{\Theta_\varrho(\varepsilon)}\|(I - E_\varepsilon)\varrho(A)Ax\| \\ &\leq \|E_\varepsilon x\| + \frac{1}{\Theta_\varrho(\varepsilon)}\|\varrho(A)Ax\|.\end{aligned}$$

Under the smallness assumption we have that

$$\|\varrho(A)Ax\| \leq \|\varrho(A)(Ax - z^\delta)\| + \|\varrho(A)z^\delta\| \leq (1 + C)\delta,$$

and also under smoothness assumption for x that $\|E_\varepsilon x\| \leq \psi(\varepsilon)$. By letting $\varepsilon_* := \Theta^{-1}_{\varrho\psi}(\delta)$ we complete the proof. □

Remark 4.7. The above result holds in much greater generality. We explain this by using the *modulus of continuity* of the (conditionally stable) Moore–Penrose inverse A^+ of A as introduced in §2.1. It is well known, that for symmetric and convex sets M this quantity provides a lower bound for any reconstruction, see [1]. In addition, for ellipsoidal sets (images of balls of linear operators in Hilbert space) this is attained, cf. [9] for the original paper, and [4] for a recent treatment. By similar reasoning as above we obtain for every centrally symmetric convex set $M \subset X$ that

$$\|x - 0\| = \|x\| \leq \omega_{\varrho}(A^{+}, M, (1 + C)\delta) \leq (1 + C)\omega_{\varrho}(A^{+}, M, \delta).$$

This even gives optimality up to a factor $1 + C$ beyond smoothness classes H_{ψ}, see Fact 1.

Proof of the main result. If there is no immediate stop ($k_{\mathrm{DP}} \geq 1$), then the bounds from Proposition 4.3 and Lemma 4.4 allow to prove the error bound. Indeed, then we have that $d_k \leq \tau\delta$, bounding the first term in the error decomposition. Lemma 4.4 provides us with

$$\alpha_k \geq \frac{\Theta_{\varrho\psi}^{-1}(\zeta\delta)}{C},$$

with $C \geq 1$ and $\zeta \leq 1$ only depending on (μ, m, τ). This implies, for some positive constant $c(\mu, m, \tau)$, that

$$\Theta_{\varrho}(\alpha_k) \geq c(\mu, m, \tau)\Theta_{\varrho}(\Theta_{\varrho\psi}^{-1}(\zeta\delta)),$$

by using that $\Theta_{\varrho} \prec \lambda^1$, see §3.1. This allows to bound

$$\begin{aligned}
\frac{1}{\Theta_{\varrho}(\alpha_k)}\delta &\leq \frac{1}{c(\mu, m, \tau)} \frac{\delta}{\Theta_{\varrho}(\Theta_{\varrho\psi}^{-1}(\zeta\delta))} \\
&= \frac{1}{\zeta c(\mu, m, \tau)} \psi(\Theta_{\varrho\psi}^{-1}(\zeta\delta)) \\
&\leq \frac{1}{\zeta c(\mu, m, \tau)} \psi(\Theta_{\varrho\psi}^{-1}(\delta)),
\end{aligned}$$

proving the assertion in this case. On the other hand, if $k_{\mathrm{DP}} = 0$, Lemma 4.6 applies (with $C := \tau$), and completes the proof in this case. □

Acknowledgments. Part of the material was presented at the meeting the IP-TA 2010 workshop *Inverse Problems: developments in theory and applications* Warsaw, February 9–12, 2010. The second author thanks the organizers for their kind hospitality.

Bibliography

[1] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of inverse problems*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[2] M. Hanke, *Conjugate gradient type methods for ill-posed problems*. Longman Scientific & Technical, Harlow, 1995.

[3] M. R. Hestenes, Relative self-adjoint operators in Hilbert space. *Pacific J. Math.* **11** (1961), 1315–1357.

[4] B. Hofmann, P. Mathé and M. Schieck, Modulus of continuity for conditionally stable ill-posed problems in Hilbert space. *J. Inverse Ill-Posed Probl.* **16** (2008), 567–585.

[5] B. Hofmann, P. Mathé and H. von Weizsäcker, Regularization in Hilbert space under unbounded operators and general source conditions. *Inverse Problems* **25** (2009), 115013, 15.

[6] P. Mathé and B. Hofmann, How general are general source conditions? *Inverse Problems* **24** (2008), 015009, 5.

[7] P. Mathé and U. Tautenhahn, *Regularization under general noise assumptions*, submitted, 2010.

[8] P. Mathé and U. Tautenhahn, *Enhancing linear regularization to treat large noise*, work in progress, 2010.

[9] A. A. Melkman and C. A. Micchelli, Optimal estimation of linear operators in Hilbert spaces from inaccurate data. *SIAM J. Numer. Anal.* **16** (1979), 87–105.

[10] V. A. Morozov, Regulyarizatsiya pri bol′shih pomehah (Russian), Regularization under large noise. *Zh. Vychisl. Mat. i Mat. Fiz.* **36** (1996), 13–21.

[11] A. S. Nemirovskiĭ, Regularizing properties of the conjugate gradient method in ill-posed problems. *Zh. Vychisl. Mat. i Mat. Fiz.* **26** (1986), 332–347, 477.

[12] R. Plato, Optimal algorithms for linear ill-posed problems yield regularization methods. *Numer. Funct. Anal. Optim.* **11** (1990), 111–118.

[13] U. Tautenhahn, Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.* **33** (1996), 2120–2130.

Received July 9, 2010.

Author information

Gilles Blanchard, Universität Potsdam, Institut für Mathematik, Am Neuen Palais 10, 14469 Potsdam, Germany.
E-mail: gilles.blanchard@wias-berlin.de

Peter Mathé, Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany.
E-mail: mathe@wias-berlin.de